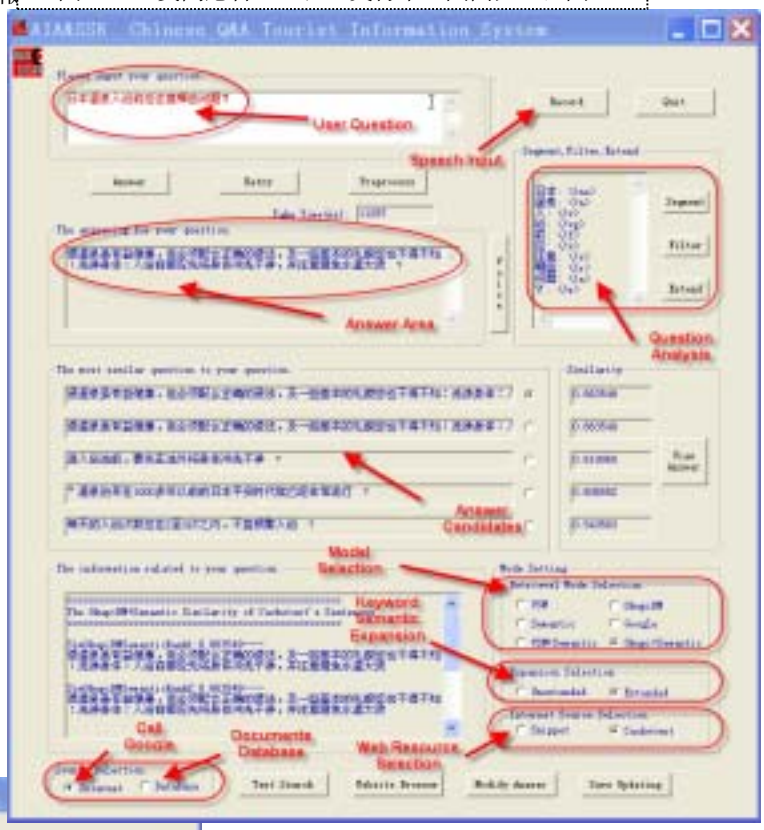
会話グループ

(1) 統合検索メソッドを用いた質問応答システムに関する研究

近年，電子化された膨大な情報源から必要な情報を効率的に獲得するため，ユーザからの自然文で問われた質問に対して､明確な回答を自動的に大量の文書から抽出する質問応答(QA :Question Answering)システムの研究が数多く行われ注目されている｡欧米や日本と比べて中国語の QA システムの研究は大変遅れていたが､近年では自然言語による中国語の QA システムについての研究が注目されるようになってきた｡効果的な QA システムは利用者の質問に的確に答えることを目指しているが,そのためには情報検索と情報

図 2-1　質問応答システム実行中の画面ダンプ図



本研究では, 回答精度及び頑健性の向上を目指し､ドメインデータベースの情報を効率的に利用できるように､統計ベースと意味解析ベースを統合した Q&A システムの構成を提案する｡オープンドメイン QA(質問の対象領域を限定しない)に対して、対象分野を限定することによって，ドメイン知識(オントロジー)の利用が容易となる。そして、より高度な言語処理を行うことで，精度の高い,実用的な QA システムの構築が可能になると期待される｡本研究では、まず、中国語の言語特徴分析と VSM モデルに基づき, よく聞かれる質問とその回答をデータベース化した質問応答データベースの回答検索と, 特定ドメイン文書の回答検索を統合した Q&A システムを構築する｡特に文書検索においては, 文書検索と文検索を統合することで，短い文で的確な回答を行うことを目指す。本研究では, ドメインを観光情報という中規模の領域に限定する。また、インターフェースにおいて従来のテキスト入出力だけでなく、音声入出力も検討する。具体的には、中国語の文型の特徴に基づいた言語モデル FSN、音響モデル HMM と発話辞書を用いて、特定領域向けの音声インターフェースとする。

図 2-2　回答検索実行中の GUI と対応するウェブサイト情報のブラウザー図

本研究では、ドメインデータベースの制限性を考慮しインターネット上のウェブ情報を利用できるように、ウェブベースの QA システムの新しいアプローチを提案した。具体的に、質問応答データベース、ドメインドキュメントデータベースとウェブリソース等の 3 種類のリソースを利用する検索メカニズムを構成し、確率モデル OkapiBM25 と HowNet 知識ベース（オントロジー）に基づく意味解析を統合する新しい検索テクニックを提案する。特に、HowNet と予め作成した特定ドメイン HowNet に基づき意味類似度と統合類似度の概念と計算手法を検討する。また、単語と文の類似度によりユーザの質問文を拡張する新しいアイデアを提案し、及びそれを利用し 3 種類のリソースデータベースを効率的に統合する手法を示す。

本研究では､提案手法の有効性とそれを用いた特定ドメイン向けの QA システムの構築可能性が検証するために、質問応答データベース検索、ドメインドキュメントデータベース検索とウェブリソース検索という三つのプロセスでいくつかの実験を行う。そして、ユーザ質問文向けのキーワード拡張のパフォーマンスを評価する。また、質問タイプにより回答精度の違いを比較し評価する。評価実験を行

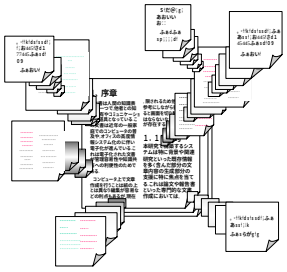
った結果、すべての質問タイプに対し提案手法により文のレベルで、平均で MRR（Mean Reciprocal Rank）が 0.8 を上回った。Factoid 型の質問を評価するのに NER（Named Entity Recognition）の処理を実行し、抽出された回答は 0.78 の MRR が達成できた。

図１、２にシステムの実行図と対応するウェブサイド情報のブラウザー図を示す。

(2)

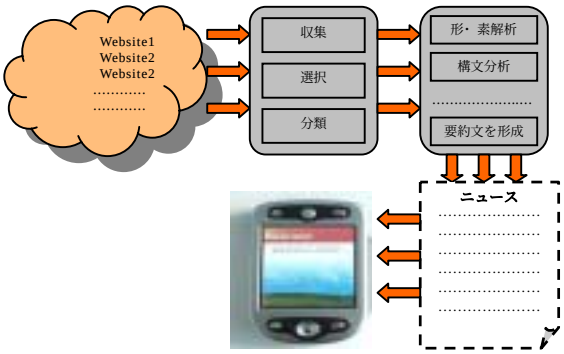
近年の中国の国際化と国々の間の交流機会増大に伴い、中国に関心をよせる外国人が増大している。現在世の中に流通する情報の圧倒的多数はテキストメディアの形態をとっている。そのため、ある国について情報を得たい場合に、最も有効な手段はその国の言葉で書かれたテキストを読むことである。それで、中国語読解支援を受けたいというニーズが増えている。特に場所や時間の制約を受けず、効果的な中国語読解支援環境の整備が強く望まれている。我々は、インターネットに接続した WWW によって、外国人に対して一つ重要な読解困難点である中国語慣用表現の読解支援システムを開発している。システムは、中国語テキストに CSL(Chinese as Second Language)学習者の中国語慣用表現の識別を支援するため、中国語慣用表現の構造と意味の特徴を充分に考慮する自然言語処理手法を利用して慣用表現の抽出を行なう。そして、認知心理学や第二言語習得論などに基づいて効果的な読解支援方法を提案する。

(3) 文章作成支援システムに関する研究

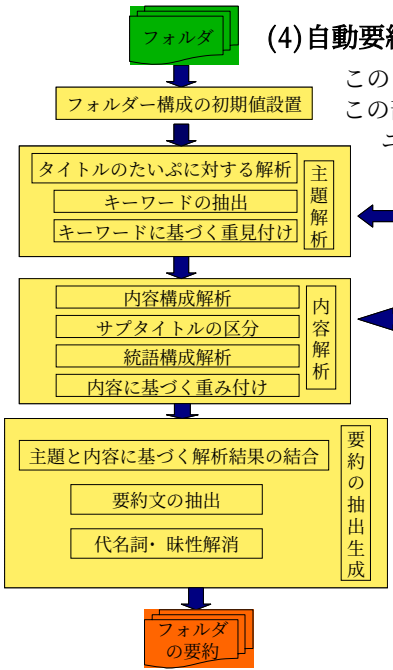


どれだけワープロが発達しても文章を書くって大変ですよね。頭の中では解っていても言葉が出てこない、論文やレポートを書いていてこんな経験がある方は多いはず。でもたくさんの論文の中をよく読んでいくと、左図みたいにある文章が他の文書の目的になっていたり、背景になっていたりと 1 つの文章が色々活用されているのに気付きます。論文を書くときも自説を補強するために背景や関連研究とか、他の人の論文を参考に文章をまとめる部分が必要なのです。そこで、この情報をまとめる部分を書く作業をお手伝いし、作者には自説の部分に力を入れてもらおうと文章の作成支援システムの構築を目指してます。まとめたい情報を複数文書要約技術を用いて自動で例文を作成し提示することで、文章作成のお手伝いします。システムのインターフェイスは下図のようになってます。言葉が出てこず苛立つ時間、減らします。

図 2-3 例文自動生成画面



(4)自動要約システム――小スクリーン設備向きの総合的な応用



このシステムは主に二つの部分からなっている。第一部分は情報の自動的サーチである。この部分は各大手ニュースホームページを情報源とし、各大手ニュースホームページのニュースを集め、選択と分類を行う。第二部分は自動要約モジュールが選択、分類済みのニュースに対して、形・素解析や構文分析や文法分析などの操作を行い、ニュースの主要的な部分を抽出し、わかりやすいニュースの要約文を形成する。また、これらの要約文は自動的にニュースホームページに発表される。インターネットと接続する移動設備のユーザーはこのホームページをアクセスすれば、ニュースの要約文を読むことができ、簡単にほしい情報を獲得できる。

主体と内容の解析に基づく自動要約

インターネットの普及、情報を得るルートの拡大とともに、毎日大量の情報は絶えずに出てくる。高速に情報の大概的な内容を理解するには役立ち、さらに全文を読むかどうか判断しやすい。自動要約技術は快速に情報を得るためのいいツールである。 本システムは部分解析に基づく方法と解析木に基づく方法を結合する方法を用いる。こうしたら、フォルダーの主題特徴と内容、構成を巧みに結合し、システムの処理性能を向上できる同時に、自動要約質も高められる。

文章作成支援システムのための素材 DB の構築について

論文や報告書の文章構成は既存情報, つまり先行研究や研究背景をまとめた部分と作者独自の研究を論ずる部分で構成されている．前者は情報の電子化などにより収集する量が増加の傾向にある．これらの情報は必要ではあるが，その論文の本質ではない．よって，作業量が増えるほど，作者の負担が大きくなると考えられる．つまり既存情報をまとめる過程を素早く書けることで作者の負担が減り本来の研究を論じることに集中でき, 全体としての負担も軽減できると考えられる．本研究の目的は論文における背景やテーマ，先行研究のような既存情報を素材とした DB の構築し，論文作成時の負担を減らすことである． 上田の文章作成支援システムにおける文章素材 DB の重要性を考え，ユーザーがより利用しやすく，効率の良い論文作成環境を提供する素材 DB の構築を提案するものである．

大規模コーパスに基づくオントロジーの自動構築

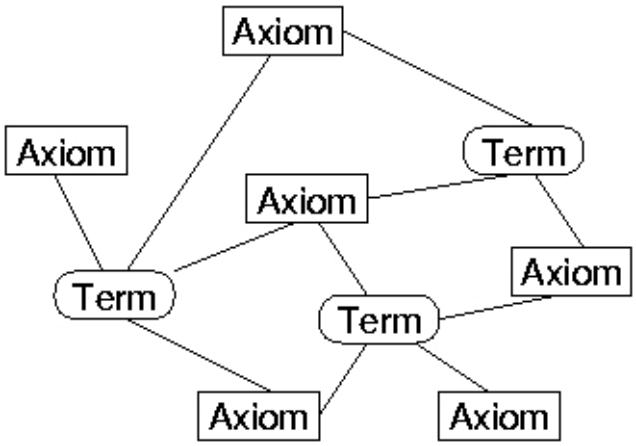
現在,情報技術の発展に伴いその技術は企業,マーケティング,研究,開発などに新しい可能性を提供しています.それに伴って,それらが扱う知識ステムの巨大化/複雑化により利用者が目的に応じた情報の収集,選択,統合を行う必要性が出てきています.現在の情報処理技術では情報を単なる記号として処理する技術として扱っていることが多く見られますが,それだけではなく,さらに知識を利用した情報処理技術として情報が表す内容を知識として扱う必要性が高まってきています． 情報が表す内容を知識として取り扱う技術としてオントロジーという技術があります.オントロジーは「知識システムを構築する際の構成要素となる基本概念/語彙の体系化」技術のことです.

しかしオントロジーの構築には人の手で構築することが一般的であり,構築には膨大な時間と専門知識が必要となることから,オントロジーの自動構築が重要な問題と考えられます..

オントロジーの構築が自動化することができれば容易に知識ベースを構築することができ,様々なシステムに応用されると期待されます..

オントロジーが用いられると考えられるツール

- 文章要約
- 情報検索（文章の意味を用いた検索など）
- 知識ベースの自動構築



図：オントロジー基本形
（Axiom：公理　Term：語彙）

動詞格フレーム情報を用いた案内図生成に関する研究

近年，音声認識や音声合成技術の発達によって，対話ロボットや音声入力カーナビゲーション等の自然言語処理技術を応用したものが多く見られるようになってきました．しかし，これらに実装されている対話機能は簡易的なものばかりであり，まだまだ実用的であるとは言えません．自然言語は人間がコミュニケーションをとる手段として昔より用いられてきたということから，それらロボットに高い言語理解能力を持たせることは，人間にとって最も親しみやすいシステムエージェントの実現につながることが期待できます．そこで我々は，人間による道案内を理解することによって自ら地図を生成できる移動ロボットの実現を目指し，案内文より案内図生成を行うシステムの構築を行っています．図 2-2 にシステムを示す．

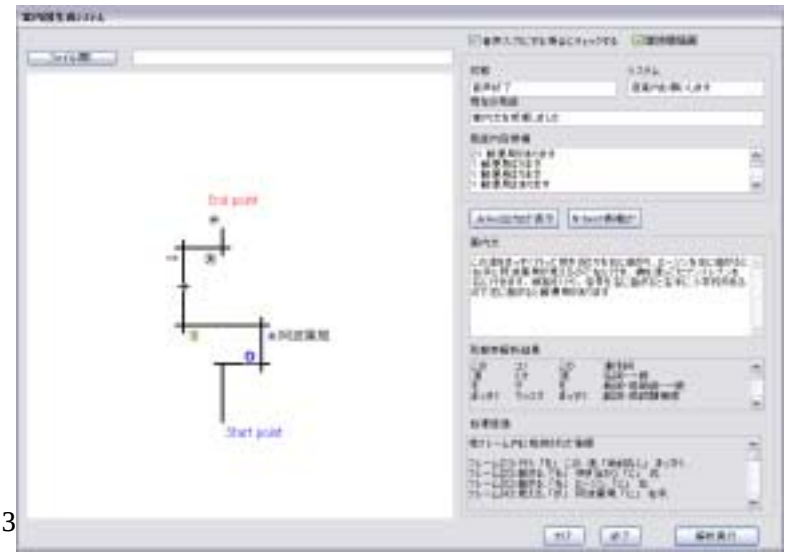


図 2-2：案内図自動生成システムの画面ダンプ図

（3）自然言語処理技術を用いた中学理科教授学習システムの作成

現在の教育に目を向けてみると，若者の理科系離れが社会的な問題となっている．これは学習の形態が知識詰め込み型，暗記中心型となり，自分で考えるという機会が少なくなったためと考えられている．理想の学習形態は，生徒ごとに専属の教育者が担当することであるが，このことが無理であることは明白である．このような背景から，コンピュータを用いた学習ソフトや教授学習システムの需要が高まっている．

学習者の理解の状況により，適切な教授を行うといった動的な環境を与えること，自然言語特有のユーモアやゲーム性を活かして，興味を引き出すことにより，飽きがこなく，繰り返し学習ができるシステムを開発することを目的とする．従来の教授学習システムでは，解答方式は選択式のものが多く学習者主体の学習環境とはいえない．本システムでは解答方式は学習者の自由文入力を可能とし，間違った解答に対してはその間違いによる現実世界での現象を提示する．本システムは，機械翻訳の分野で提案された Super Function を用いて，学習者により入力された文章の解析，理科問題文の自動生成を行う．またシステムのインターフェースには音声認識，音声合成を用いる．図 2-3，2-4 にシステムのインターフェース画面を示す．

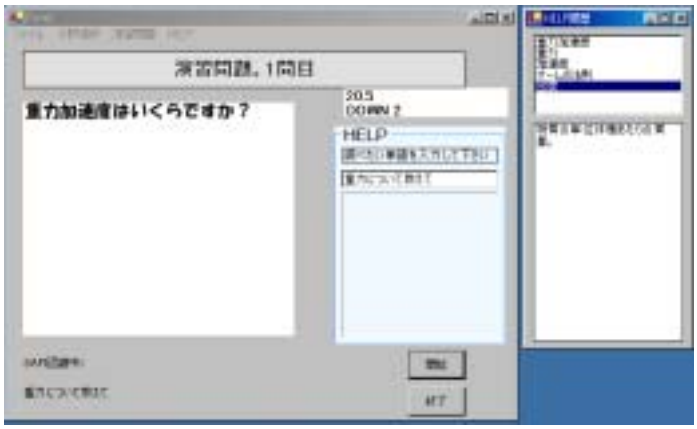


図 2-3：演習問題解答中のヘルプ表示

図 2-4：理科問題文入力

会話システムの知識ベース構築について

近年，インターネットの普及によってさまざまな情報が入手できるようになっています．また，百科事典や新聞を始め，さまざまな文字情報の電子化が進み，それらの活用は日常の中で不可欠とも言える要素となっています．しかし，個人がアクセス可能な情報は膨大で，必要な情報を効率よく得るのは困難です．また，検索単語によっては不必要な情報が膨大に検出されることもあり，必要な情報を得るまでに時間が掛かることもあるのが，現状です．

膨大な情報の中から必要な情報を効率よく取り出す方法の一つとして，質問応答システムを始めとする会話システムが注目されています．ここでいう質問応答システムとは，ユーザによる自然言語での質問に対し，適切な回答を返すシステムを指します．会話システムを構成する上で，重要な課題となってくるのが知識ベースの構築です．インターネット上に既に存在する電子化された文書を利用した質問応答システムの研究も，盛んになされています．また，情報の範囲を限定し，知識ベースを人手で作成した自動案内システムなども考案されています．この知識ベースの構成や情報量によって，システムの回答範囲が決定すると言えます．本研究では，PtoPAのソフトCAIWA2.1に基づいて知識ベースを構築すると共に，現在の会話システムの問題点，また，今後必要と思われる能力について提案を行います．作成した知識ベースを用いて実際に会話を行っている画面を図に示す．